

# Williams College Archives and Special Collections File Format Recommendations

Based on current digital data curation best practice and recommendations, the Williams College Archives has set the following confidence levels for the longevity of commonly used file formats. Record creators are encouraged to take these recommendations into consideration when determining what file formats to save records as during their active lifecycle. The longer the active lifecycle is anticipated to be, the greater the need to use formats ranked as medium or high confidence.

Media type	High Confidence Level	Medium Confidence Level	Low Confidence Level
Text	<ul style="list-style-type: none"> <li>- Plain text (encoding: USASCII, UTF-8, UTF-16 with BOM)</li> <li>- XML (includes XSD/XSL/XHTML, etc.; with included or accessible schema and character encoding explicitly specified)</li> <li>- PDF/A-1 (ISO 19005-1) (.pdf)</li> </ul>	<ul style="list-style-type: none"> <li>- Cascading Style Sheets (.css)</li> <li>- DTD (.dtd)</li> <li>- Plain text (ISO 8859-1 encoding)</li> <li>- PDF (.pdf) (embedded fonts)</li> <li>- Rich Text Format 1.x (.rtf)</li> <li>- HTML (include a DOCTYPE declaration)</li> <li>- SGML (.sgml)</li> <li>- Open Office (.sxw/.odt)</li> <li>- OOXML (ISO/IEC DIS 29500) (.docx)</li> </ul>	<ul style="list-style-type: none"> <li>- PDF (.pdf) (encrypted)</li> <li>- Microsoft Word (.doc)</li> <li>- WordPerfect (.wpd)</li> <li>- DVI (.dvi)</li> <li>- All other text formats not listed here</li> </ul>
Raster Images	<ul style="list-style-type: none"> <li>- TIFF (uncompressed)</li> <li>- JPEG2000 (lossless) (.jp2)</li> <li>- PNG</li> </ul>	<ul style="list-style-type: none"> <li>- BMP (.bmp)</li> <li>- JPEG/JFIF (.jpg)</li> <li>- JPEG2000 (lossy) (.jp2)</li> <li>- TIFF (compressed)</li> <li>- GIF (.gif)</li> <li>- Digital Negative DNG (.dng)</li> <li>- PNG (.png)</li> </ul>	<ul style="list-style-type: none"> <li>- MrSID (.sid)</li> <li>- TIFF (in Planar format)</li> <li>- FlashPix (.fpx)</li> <li>- PhotoShop (.psd)</li> <li>- RAW</li> <li>- JPEG 2000 Part 2 (*.jpf, .jpx)</li> <li>- All other raster image formats not listed here</li> </ul>
Vector Graphics	<ul style="list-style-type: none"> <li>- SVG (no Java script binding) (.svg)</li> </ul>	<ul style="list-style-type: none"> <li>- Computer Graphic Metafile (CGM, WebCGM) (.cgm)</li> </ul>	<ul style="list-style-type: none"> <li>- Encapsulated Postscript (EPS)</li> <li>- Macromedia Flash (.swf)</li> <li>- All other vector image formats not listed here</li> </ul>
Audio	<ul style="list-style-type: none"> <li>- AIFF (PCM) (.aif, .aiff)</li> <li>- WAV (PCM) (.wav)</li> </ul>	<ul style="list-style-type: none"> <li>- SUN Audio (uncompressed)</li> </ul>	<ul style="list-style-type: none"> <li>- AIFC (compressed) (.aifc)</li> </ul>

		<ul style="list-style-type: none"> <li>(.au)</li> <li>- Standard MIDI (.mid, .midi)</li> <li>- Ogg Vorbis (.ogg)</li> <li>- Free Lossless Audio Codec (.flac)</li> <li>- Advance Audio Coding (.mp4, .m4a, .aac)</li> <li>- MP3 (MPEG-1/2, Layer 3) (.mp3)</li> </ul>	<ul style="list-style-type: none"> <li>- NeXT SND (.snd)</li> <li>- RealNetworks 'Real Audio' (.ra, rm, .ram)</li> <li>- Windows Media Audio (.wma)</li> <li>- Protected AAC (.m4p)</li> <li>- WAV (compressed) (.wav)</li> <li>- All other audio formats not listed here</li> </ul>
Video	<ul style="list-style-type: none"> <li>- Motion JPEG 2000 (ISO/IEC 15444-4) (.mj2)</li> <li>- AVI (uncompressed, motion JPEG) (.avi)</li> <li>- QuickTime Movie (uncompressed, motion JPEG) (.mov)</li> </ul>	<ul style="list-style-type: none"> <li>- Ogg Theora (.ogg)</li> <li>- MPEG-1, MPEG-2 (.mpg, .mpeg, wrapped in AVI, MOV)</li> <li>- MPEG-4 (H.263, H.264) (.mp4, wrapped in AVI, MOV)</li> </ul>	<ul style="list-style-type: none"> <li>- AVI (others) (.avi)</li> <li>- QuickTime Movie (others) (.mov)</li> <li>- RealNetworks 'Real Video' (.rv)</li> <li>- Windows Media Video (.wmv)</li> <li>- All other video formats not listed here</li> </ul>
Spreadsheet/ Database	<ul style="list-style-type: none"> <li>- Comma Separated Values (.csv)</li> <li>- Delimited Text (.txt)</li> <li>- SQL DDL</li> </ul>	<ul style="list-style-type: none"> <li>- DBF (.dbf)</li> <li>- OpenOffice (.sxc/.ods)</li> <li>- OOXML (ISO/IEC DIS 29500) (.xlsx)</li> </ul>	<ul style="list-style-type: none"> <li>- Excel (.xls)</li> <li>- All other spreadsheet/ database formats not listed here</li> </ul>
Virtual Reality	<ul style="list-style-type: none"> <li>- X3D (.x3d)</li> </ul>	<ul style="list-style-type: none"> <li>- VRML (.wrl, .vrmf)</li> <li>- U3D (Universal 3D file format)</li> </ul>	<ul style="list-style-type: none"> <li>- All other virtual reality formats not listed here</li> </ul>
Computer Programs		<ul style="list-style-type: none"> <li>- Computer program source code (.c, .c++, .java, .js, .jsp, .php, .pl, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>- Compiled / Executable files (EXE, .class, COM, DLL, BIN, DRV, OVL, SYS, PIF)</li> </ul>
Presentation Files		<ul style="list-style-type: none"> <li>- OpenOffice (.sxi/.odp)</li> <li>- OOXML (ISO/IEC DIS 29500) (.pptx)</li> </ul>	<ul style="list-style-type: none"> <li>- PowerPoint (.ppt)</li> <li>- All other presentation formats not listed here</li> </ul>

# Williams College Archives and Special Collections Media Type Preservation Plan

Once records have become inactive and are transferred to the Williams College Archives, the Archives' primary preservation strategy is to normalize files to preservation and access formats upon ingest. The choice of access formats is based on the ubiquity of viewers for the file format. All preservation formats are open standards. Additionally, the choice of preservation format is based on community best practices, availability of open-source normalization tools, and an analysis of the significant characteristics of each media type. The College Archives also maintains the original format of all ingested files to support future migration and emulation preservation strategies as needed. The normalization chart below is based on international best practice standards and the default plan for Archivematica. The Archives monitors standard format registries and technological advances and adjusts the plan as needed.

Media type	Original File formats	Preservation format(s)	Access format(s)	Normalization tool
Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Portable Document Format	PDF	PDF/A	PDF or PDF/A	Ghostscript
Presentation files	PPT	ODF	PDF	OpenOffice
Images	BMP, GIF, JPG, JP2, PNG, PSD*, TIFF, TGA	Uncompressed TIFF	JPEG	ImageMagick
Raw camera files	NEF	DNG	JPEG	DigiKam DNG Converter
Spreadsheet/ Database	XLS	ODF	Original format	Unoconv/OpenOffice
Plain text	TXT	Original format	Original format	None
Vector images	AI*, EPS*, SVG*	SVG	PDF	Inkscape
Video	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4*, SWF, WMV	MPEG-2	MPG	FFmpeg
Word processing files	DOC, WPD, RTF	ODF	PDF	OpenOffice